R E P O R T R E S U M E S

ED 014 465 SP 001 363

CRITERIA--PROBLEMS IN VALIDATING TEACHER SELECTION POLICIES AND PROCEDURES.

BY- RYANS, DAVID G.
NEW YORK CITY BOARD OF EDUCATION, BROOKLYN, N.Y.

PUB DATE JUN 67

GRANT OEG-1-6-061665-1624
EDRS PRICE MF-$0.25 HC-$0.68 15P.

DESCRIPTORS- *CRITERIA, *EFFECTIVE TEACHING, LESSON OBSERVATION CRITERIA, *PREDICTIVE MEASUREMENT, PREDICTIVE VALIDITY, TEACHER EVALUATION, *TEACHER SELECTION, *TEACHING VALUES, VALUES,

THE VARIOUS WEAKNESSES, DIFFICULTIES, AND "COMMON CONFUSIONS" THAT HAVE CHARACTERIZED THE CRITERIA USED IN TEACHER SELECTION PROCEDURES ARE IDENTIFIED, AND SUGGESTIONS FOR IMPROVEMENT ARE MADE. TEN STEPS IN THE DEVELOPMENT AND EVALUATION OF PROCEDURES ARE OUTLINED--FROM IDENTIFICATION OF LOCALLY VARYING EXPECTATIONS ARISING FROM VARYING VALUES (E.G., IS THE TEACHER TO BE PERMISSIVE OR A DISCIPLINARIAN) THROUGH DRAWING INFERENCES ABOUT THE VALIDITY OF THE PROCEDURES FOR PREDICTING THE (OPERATIONALLY DEFINED) CRITERION BEHAVIORS. CRITERIA MUST BE CLEARLY DISTINGUISHED FROM CRITERION MEASURES. PROBLEMS OF THE VALIDITY AND GENERALIZABILITY OF CRITERION DESCRIPTIONS AND OF THE VALIDITY AND REALIABILITY OF PROCEDURES FOR ESTIMATING CRITERION BEHAVIORS ARE DISCUSSED (E.G., IS THE BEHAVIOR UNI- OR MULTIDIMENSIONAL. ARE THE DIMENSIONS DISCRETE, REPRESENTATIVE, GENERALIZABLE, REPLICABLE). A VARIETY OF APPROACHES TO JUDGING THE VALIDITY OF CRITERION ESTIMATES AND TO DESCRIBING CRITERIA ARE GIVEN, AND SIX YARDSTICKS FOR JUDGING THE USEFULNESS OF A CRITERION DESCRIPTION ARE OUTLINED. THE MOST VALID OF SEVERAL METHODS OF OBTAINING CRITERION ESTIMATES USES SAMPLES OF THE ON-GOING CRITERION BEHAVIORS AND DIRECT ESTIMATION BASED ON OBSERVATION OF THEM IN PROCESS. THIS DOCUMENT APPEARED IN GILBERT, H.B., AND LANG, G., "TEACHER SELECTION METHODS," NEW YORK, 1967. (RP)

SP 001363

ED014465

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TEACHER SELECTION METHODS

Project No. 6-1665
Grant No. OEG 1-6-061665-1624

Harry B. Gilbert

Pennsylvania State University

Gerhard Lang

Montclair State College

June 1967

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

BOARD OF EXAMINERS
Board of Education of
The City of New York

New York, New York

Criteria: Problems in Validating Teacher Selection Policies and Procedures

David G. Ryans

University of Hawaii

Nature of Criteria

A good deal of confusion appears to exist with regard to just what we mean when the term criterion, or criteria, is employed. A criterion is simply a standard or bench-mark used to provide a frame of reference for judging or evaluating something. It may be thought of as a model against which comparisons may be made. Usually criteria evolve from common agreement about acceptable standards--regulatory boards for insurance, public utilities, banking, contracting, and such operate with a set of agreed-upon standards (criteria) as a model. In many circumstances criteria are arbitrary and relative to values that are held to be important by some particular group of persons at some particular time and place. Indeed, the matter of "values" and "value systems" is basic to the consideration of criteria.

In relation to teacher selection, just as opinions and preferences (values) of individuals vary with regard to the competencies and behaviors expected of teachers, the criteria against which teacher selection procedures should be compared will often vary (at least, in certain features) from community to community; and validity studies of teacher selection nearly always require replication adapted to the varying conditions.

In taking this position that criteria are determined by value contexts that differ among schools and communities, I am implying that the first step in the consideration of criteria against which to judge a teacher selection program must be to determine the expectations that are held locally with regard to teaching and teacher behavior. The extent to which there may be consensus about major issues, the greater the assurance with which a school administration or its educational researchers may approach the designation of criteria and their components, and the greater the possibility of conducting meaningful validity studies of teacher selection procedures.

I shall return to the relationship between criteria and value systems later, but I first would like to comment further on the nature of criteria--and the frequent neglect of and confusions about considerations relating to criteria.

Here I do not restrict my remarks to studies of teacher selection and research on teacher behavior; they are no more vulnerable than a great deal of research in the behavioral sciences where the problem of the dependent variable, or the criterion, has been neglected. From years of reading research reports and research proposals I conclude that many otherwise elegantly designed researches--well designed from the standpoint of sampling, control of the experimental variable and other independent variables, data analysis (and often involving real ingenuity of approach)--have, almost as an after-thought it seems settled upon some available instrument, or perhaps hastily thrown together some test, inventory, or other observational technique without great regard to its validity and reliability, proceeding on the assumption that such instrument satisfactorily reflected, and provided useful estimates of, the criterion behavior. This happens to be a pet peeve of mine--the fact that so many investigators seem to neglect or give too little attention to control of the dependent variable (criterion behavior) and that instead of considering this important problem from the very beginning of their research they appear to give it only cursory attention some time later in the investigation. I suspect in many such cases the researcher is introducing a source of Type II error; or when statistically significant relationships between experimental variables and the assumed dependent variable are obtained, the relationship really may be between the experimental variable and an unintentionally biased and unsatisfactory estimate of the criterion behavior. Although I feel my accusation is fairly generally applicable to research, it is one about which we teacher selection researchers certainly need to do some considered soul-searching.

Teacher Selection in Perspective

It seems to me there is some similarity between what we are interested in when we plan teacher selection procedures and subsequently study their usefulness and the sorts of things a curriculum developer is concerned with.

Typically, I believe, the planner of instructional techniques and course materials considers the ideal procedure to be followed as consisting of: (1) designation of course objectives, goals, and expectations; (2) breaking down the objectives into descriptions of (a) expected teacher behaviors and (b) expected pupil behaviors, i.e., the pupil behaviors it is intended the course or curriculum will help to nurture and develop; (3) planning and development of specific curricular materials and instructional techniques that are hypothesized to aid in developing the intended pupil behaviors; (4) selection of appropriate means of measuring the attainment of the behaviorally described objectives (Recall that any

one of a number of methods may be used--e.g., measurement of samples of the pupil criterion behavior (samples of the expected pupil behaviors), measurement of aspects of teacher performance, measurement of teacher opinion about the efficacy of the program, citing of critical incidents, measurement of pupil behavior known to be related to the criterion behavior, measurement of pupil test response to verbally described situations related to the criterion behavior, etc.); (5) assembly of data (which may be in any of several kinds of units, scores, ratings, etc.) yielded by the measurement devices that were assumed to reflect attainment of the specified objectives; and (6) evaluation of the course materials and/or instructional procedures by drawing inferences from the collected data about attainment of the course objectives.

It may be laboring the obvious to spell out the closely related steps that are involved in the development of teacher selection procedures and their evaluation. Nevertheless I am going to describe what I believe to be a procedure that provides an appropriate rationale from which teacher selection should proceed. (Note that I consider this procedure to represent an "ideal" one--one which often cannot be followed step-by-step in practice. Practical considerations often demand that we skip early phases and proceed to set up teacher selection techniques on the basis of best available judgment--and I should not imply the selection procedures thus developed necessarily will be poor; they may be based upon substantial wisdom growing from experience, and upon testing them out they may be found to yield results that can indeed be shown to relate to valid criteria of teacher behavior, even though these criteria were not determined prior to the planning of the teaching program. I think we are getting the cart before the horse to develop selection procedures and then at some later time turn attention to the criteria to which we think these procedures ought to relate--but sometimes this is the best we can do.)

But let me get on to a statement of what I think we might agree would be a desirable way to proceed if we were operating within an ideal situation. It is a procedure that is fairly similar to the curriculum development paradigm I spelled-out a moment ago. I will refer to some ten steps or phases.

> (1) Selection and designation of general aspects of the value system framework of the school/community as they relate to teacher behavior. I am referring here to the agreed-upon qualities that are desired, or expected, of teachers in a particular place and in particular kinds of teaching situations. (Note again that this process of arriving at criteria necessarily is subjective and a matter of the values individuals or groups of individuals may possess in common. When we designate criteria we

proceed from a context of an accepted value system. We view teacher behavior in light of a set of attitudes, opinions, and viewpoints that reflect the sorts of teacher behavior we approve and prefer and also the kinds of behavior we disapprove and find unacceptable. Value judgments and the value concepts and systems on which they are based grow out of highly personal biases, preferences, beliefs, opinions, and attitudes we hold as individuals. To the extent any group of persons share in common certain expectancies, preferences, or biases about teachers and teaching, criteria of teacher behavior may be defined for that particular group. Thus value systems concerning teaching, and criteria of teacher behavior, are likely to be relative rather than absolute. Although some "valued teacher behaviors" may be held in common by a large cross section of citizens and educators at a particular time, still other "valued behaviors" that must be taken into account in specifying criteria may vary from one community or school to another.)

(2) Identification of kinds of situations in which the agreed-upon "valued teacher behaviors" may occur--and in which they may be observed and assessed.

(3) Operational description (i.e., description in terms of actual teacher behaviors) of the agreed-upon valued behaviors that are to comprise the criteria of teacher behavior.

(4) Selection of methods of estimating the operationally (i.e., behaviorally) described valued behaviors. This is the problem of instrumentation relative to the criterion behavior and obtaining assessments of the criterion behaviors. (Assessment relates to quantified, or quasi-quantified, description. When we make an assessment of some characteristic of some thing or some behavior, we are concerned with the degree to which that characteristic is manifest.) In assessing some aspect or characteristic of the criterion behavior of teachers we are trying to estimate the extent to which that defined characteristic is manifest by some teacher.

(5) Identification of observable properties of teacher classroom behavior that may be related to the specified operationally described criteria (i.e., the descriptive cataloguing of teacher characteristics and behaviors that occur in the classroom).

(6) Development of selection instruments and procedures that are hypothesized to yield estimates that will reflect the operationally described teacher behaviors (criterion behaviors)--which, in turn, are assumed to reflect the value framework of the school and the community served.

(7) Assembly of data yielded by the teacher selection instruments and procedures noted in Step 6 above.

(8) Assembly of data yielded by the procedures used to estimate the criterion behaviors--Step 4 above.

(9) Analysis of relationships between estimates of the behaviorally defined criterion behavior and the estimates of teacher characteristics used in the teacher selection procedure.

(10) Evaluation of the teacher selection procedures by the drawing of inferences about the validity of those procedures for predicting the criterion behaviors designated in Steps 1 through 3 above.

Common Confusions in Dealing with Aspects of the Criterion Problem

One of the reasons we have difficulty with the criterion problem is that we sometimes fail to distinguish between different aspects of what is involved. We have all heard "principals' ratings" referred to as a criterion of teacher behavior. I think it is not nit-picking to note that principals' ratings do not constitute a criterion or a description of a criterion behavior. They are one kind of estimate, derived from one method of obtaining data that may, under some conditions perhaps, be related to some specified aspect of the criterion behavior of teachers.

Allow me to illustrate what I mean about confusion of terms with one or two examples.

Let us suppose the value system in a particular school community expects its teachers to possess some degree of capability with respect to "classroom management." We may think of this, I believe, as a criterion of teacher behavior in that community. It is, of course, a very generalized and abstract description of criterion behavior at this stage. Before we can proceed to observe teachers with respect to their capabilities for classroom management we need to specify still further the kinds of behaviors that comprise this domain and we need to try to determine either (a) samples of the criterion behavior that may in some manner be observed

and assessed and/or (b) known or assumed correlates of this criterion behavior that may be observed and assessed. As one example, of many that could be used, we might choose the teacher's response to a situation involving activities on the part of some pupil that interfered with the activities of his classmates in pursuit of the objectives of instruction. We are still talking about criterion behavior but we now have broken it down into a description (although still somewhat general) of a sample of the criterion behavior. Now we might choose any one of several methods of estimating the criterion behavior under consideration. And the method we would use would determine, at least to a large extent, the kinds of assessments or estimates of the specified criterion behavior we would obtain. We might choose to employ direct observation of teachers in the classroom by trained observers, and one of the kinds of estimates we might obtain by such a method would be observers' ratings recorded on a scale representing qualitatively defined degrees of appropriate teacher behavior in the disciplinary situation referred to. Or, we might choose to use principals' recall of teacher behavior in situations involving classroom management and this method might yield estimates in terms of some sort of ratings, rankings, etc.

As a second possible example, suppose it was agreed that one aspect of the criterion behavior of a teacher should be his capability of communication of knowledge. In our attempt at behavioral description of the criterion, one aspect of communication of knowledge might be determined to be the teacher's behavior in a situation involving the presentation and explanation of specified subject matter content. (This could be made still more specific--we might specify behavior that emphasized clarity and directness of presentation, or perhaps subject matter depth, or absence of irrelevancies, etc.) In light of such criterion behavior we might resort to an observation method that involved the use of teacher examinations which would yield estimates of the teacher's knowledge and understanding of the specified subject matter. Or, we might again resort to direct observation by trained observers and obtain ratings, frequency counts, or other kinds of estimates. Or, instead of employing a sample of the criterion behavior of the teacher per se, we might choose to view the criterion in terms of known or assumed correlates of "teacher communication of knowledge." In this case we might choose to measure pupil knowledge of particular facts, principles, etc. that are assumed to be a product (at least in part) of the teacher's behavior in the communication of the specified knowledge. In this case as a method of estimating the criterion behavior we might elect to test pupil knowledge before the teacher presentation and immediately after presentation, obtaining estimates of the differences in test-estimated pupil knowledge before and after exposure to the teacher's presentation;

or we might test pupil knowledge before the teacher presentation and again after some specified period of time--to obtain estimates of the extent to which the teacher communicated knowledge was retained by the pupil; or we might use a method of determining the success of the pupil in later situations for which the specified knowledge is presumed to be a necessary prerequisite--such procedure yielding estimates based on test scores and grades in subsequent units of a course, in advanced courses, etc.

I have used these examples to help distinguish aspects of the criterion problem that sometimes are confused when we discuss such matters. The methods of estimation and the estimates yielded by different methods of estimating criterion behavior should, I think, be clearly distinguished in our thinking from the descriptions of criteria against which teacher selection procedures may be evaluated. The criteria themselves are the behaviors of teachers that are held to be of value. And of particular importance to validity studies of selection procedures, we need to recognize identifiable behavior samples and known correlates of the "valued" behaviors that are accepted as the criteria of teacher competency.

Some Considerations in the Designation and Estimation of Criteria

I would like now to note, at least in outline, some of the kinds of problems we must face in dealing with criteria. I will restrict my comments to two types of problems. One set of concerns has to do with (a) the validity of the description of samples and/or correlates of criterion behavior (i.e., the validity of criterion descriptions in light of the value system involved) and (b) the generalizability of descriptions of criterion behavior. The other set of problems has to do with the validity and reliability of procedures that may be used for estimating specified criterion behaviors of teachers. I shall mention, but not discuss in any detail, three different concerns from the standpoint of the validity of definitions and descriptions of criterion behavior in teaching.

One such area of concern has to do with judgments about the dimensionality of the criterion behavior under consideration: (1) Is the criterion behavior uni- or multidimensional? (Needless to say we usually agree that teacher behavior involves a number of dimensions that interact in complex combinations.), (2) How do behaviors that comprise important dimensions of teacher behavior aggregate--what are the behavior aggregates or patterns that really are relevant from the standpoint of teacher classroom behavior? What is the relative importance of various dimensions of criterion behavior in teaching--and how should these be weighted in criterion description?

A second set of concerns having to do with validity of the criterion description regard the logical consistency and interrelatedness of criterion dimensions--(1) How are the component dimensions of the criterion behavior patterned? (2) How do they overlap?

Still another area of concern from the standpoint of validity of criterion description has to do with the sampling adequacy or representativeness of the criterion dimensions that are selected to reflect criterion behavior. This is essentially the problem of trying to arrive at a criterion description that is as free as possible of bias. A number of sources of criterion bias were described almost twenty years ago by Brogden and Taylor in their classic article on "The Theory and Classification of Criterion Bias" (Educational and Psychological Measurement, 1950, 10, 159-186.) I reviewed the bias problem drawing heavily upon the insightful Brogden-Taylor treatment, and other considerations of the criterion, in The Journal of Genetic Psychology in 1957 ("Notes on the Criterion Problem in Research, with Special Reference to the Study of Teacher Characteristics," J. Genetic Psychology, 1957, 91, 33-61.) Since more detailed discussions exist, I will only remind here that in designating criteria against which to judge teacher selection techniques we must know how to recognize and must be constantly aware of conditions that may bias (and make useless) our criterion descriptions. I refer particularly to contamination bias, opportunity bias, experience bias, rating bias, deficiency (incompleteness) bias and distortion bias.

With regard to the generalizability of the criterion definition and description, (and here I am speaking of the replicability of the criterion description under different circumstances) I again note the relation of criteria to value systems espoused by a group and the probable variation in adequate criterion descriptions in different communities. We need be concerned whether criterion descriptions may vary from one kind of teaching situation to another for the same teacher and from one sample of teachers to another.

Once the problem of criterion description has been faced, we must deal with considerations relative to choice of the method or methods, of estimating the criterion behavior and the kinds of measurements or estimates that may be employed. Here again we are faced with the problems of validity and the reliability--this time, the validity and reliability of the instruments and the data they yield with respect to the criterion descriptions we have selected with view to their validity or relevance.

A variety of approaches may conceivably be applied to judging the validity of criterion estimates. Often the researcher

concerns himself only with "face" validity, where the method for estimating the criterion behavior is superficially judged to be reflecting the criterion behavior it purports to measure. The behavior elicited is assumed to be isomorphic with the criterion behavior.

Sometimes an approach which I will refer to as "postulational validity" is employed. Here the method and estimates for assessing the criterion behavior are judged, in light of postulated relationship of the behavior elicited to the criterion behavior, to be measuring the criterion. Various sub-approaches to the determination of the postulational validity include: validity by definition; validity judged from the existence of reliable differences between individuals when the method is applied; content validity, or assumption of validity based upon estimates derived from selected samples of the criterion behavior; and validity in terms of conceptual consistency--validity of the estimation method judged in light of the apparent relationship between estimates provided by the method employed and some inferentially identified "construct" or behavior.

Further important considerations with regard to criterion estimates or measurements have to do with the reliability of data yielded by a particular method; and also with the feasibility, or practicability, of an estimating technique. Certainly these cannot be neglected.

Approaches to Criterion Definition and Description

Returning to the matter of descriptions of criterion behavior, I should like to simply note some of the approaches that may be employed.

Criterion description is basically a function of thorough and detailed acquaintance with the behavior we are dealing with--in this case, the behavior of teachers as they carry out the responsibilities demanded of them in particular school situations. Such acquaintance usually is best acquired by controlled observation. This is a particularly important consideration. Too often, I suspect, we try to accomplish criterion description by arm-chair and associative recall methods. The generality and usefulness of a criterion description is likely to be proportionate to the extent that essential details of the behavior under study have been identified and classified. And the most appropriate way of becoming knowledgeable about behaviors that may contribute to the criterion is by observation under controlled conditions.

Generally speaking, the usefulness of satisfactoriness of a criterion description will be greater when:

(a) the criterion behavior under consideration, or its products, can be operationally described, directly observed, and objectively recorded;

(b) the possibility of varied interpretations of the criterion behavior and its products by different individuals is minimum;

(c) the observations directed at the identification of criterion behaviors and the data based on observations are analytical rather than global;

(d) meaningful aggregates of the criterion behavior are distinguishable from irrelevant behaviors and attention is given to the determination of such behavior patterns;

(e) the investigator is cognizant of, and attentive to, the more prevalent sources of criterion bias (e.g., contamination by concomitant behaviors, by "opportunity," by experience, by rating sets, etc.; deficiency, or incompleteness of the criterion; and criterion distortion);

(f) observations directed at identification of criterion behaviors have been extensively replicated (e.g., an adequate number of individual cases have been observed and observations conducted in a variety of times, places, and circumstances).

And we must remember that underlying all criterion descriptions is the matter of identifying the prevailing values and expectations that form the context for, and dictate, the criteria we formulate. We must first seek answers to value oriented questions such as: Are teachers expected to be permissive with regard to pupil behavior? Are they expected to maintain rigorous standards of pupil learning and control? Are teachers expected to be rigid disciplinarians? Are teachers expected to be highly knowledgeable about subject matter content? Are they expected to be available repositories of subject matter knowledge, or are they expected to arrange for the pupil to "discover" information? Are they expected to take an active part in directing learning, or to arrange learning situations for individual progress? Are teachers expected to participate in administrative policies and decisions, or are such matters to be left to the administrative staff?

Obviously these are only a few questions illustrative of a kind that might be asked in trying to assess the value climate of a community or school. The questions referred to admittedly relate to global sorts of values--behavioral descriptions would have to be derived in greater specificity to be useful. Questions of this sort

do not refer to all-or-none value judgments. They are not necessarily mutually exclusive. And the answers do not spell-out the behavioral criteria to be employed in judging the validity of a teacher selection policy or procedure. Nevertheless, such questions, together with many others, do provide the necessary first step of determining the value climate before the process of designating specific criteria can be engaged-in.

The actual description and definition of criterion behavior may follow a variety of approaches or strategies. All too frequently (or so it appears at least) no strategy at all is followed, i.e., criterion definition is completely neglected, or at best given only brief attention resulting in non-critical assumption of the criterion behavior involved. Among the studies I have reviewed I find practices covering a wide range of acceptability: completely non-critical assumption of the criterion behavior (either failure to consider criterion definition or unsophisticated acceptance of a criterion definition with no attention to its (a) completeness or (b) freedom from contaminating and distorting conditions); criterion description based upon analyses of judgments of presumably qualified authorities; criterion description based upon the analysis of responses to some response-evoking technique which is hypothesized to reflect some criterion behavior (here the criterion description is derived from the method of estimating the criterion--a procedure that should give us pause); and criterion behavior identified by analysis of records based upon observation of (a) behavior in situations presumed to involve the criterion behavior or (b) products of behavior in situations presumed to involve the criterion behavior.

Approaches to Obtaining Criterion Data

Assuming we can describe our criteria satisfactorily we can now turn to ways of obtaining criterion estimates, i.e., the basic records and indices of criterion behavior against which data derived from selection procedures may be compared.

As we have noted before a variety of methods of estimating criterion behavior are available--methods which vary in rationale and also in usefulness. May I just mention some of these in outline form:

Some methods of obtaining criterion estimates

A. Obtaining samples of the criterion behavior

1. Direct measurement of samples of the criterion behavior in process (i.e., on-going behavior)--primary criterion data.

a. "Natural" behavior--i.e., uncontrolled typical behavior

b. Standard samples of the criterion behavior

(1) Direct observation and assessment of behavior (including interview) by trained observers

(a) some observation approaches

--Systematic, with immediate assessment (time sampling)
--Retrospective (nonsystematic)
--Analytical
--Global
--Relative
--Absolute

(b) procedures

--Rating devices
--Check lists

(2) Observation and assessment of preserved records of criterion behavior in process (e.g., video tapes)

(3) Assessment by untrained observers

2. Measurement of samples of <u>products</u> of the criterion behavior--presumed products of <u>primary</u> criterion data

a. Direct observation and assessment of samples of behavior <u>products</u> e.g., on-going pupil behavior

--Uncontrolled products (i.e., products in natural situations)
--Standard samples of products

b. Use of devices for immediately eliciting the products of criterion behavior

(1) estimations of maximum performance, e.g., pupil test results

(2) estimation of typical performance, e.g., pupil responses to personal reaction questionnaires (self reports of opinions, temperamental responses, etc.)

(3) Measurement of (a) <u>change in process</u>, or (b) <u>change in product</u>

(a) change in estimates of samples of directly observed on-going teacher behavior

(b) change in estimates of samples of a presumed product of the criterion behavior, i.e., pupil behavior

B. Identification of correlates of the criterion behavior (i.e., behavior "in process" or products which may be used as signs of the criterion behavior)--secondary criterion data.

In my opinion the most valid of the various methods of estimating criterion behavior is that of focusing upon samples of the on-going criterion behavior and resorting to direct estimation based on observation of these samples of criterion behavior in process.

Ideally, in the study of the validity of teacher selection procedures one would prefer to work with "identical elements" of the criterion behavior in which he is interested--to directly observe and directly measure the samples of the criterion behavior on which attention is focused. We would like to employ measurements based on "work samples" or the "natural" or "typical" behavior in process, or, as a second best choice, upon similar observations of a product of the criterion behavior. In many cases I think we can accomplish our study in this manner. In others, it is true, we must be satisfied with the indirect estimates or correlates of the criterion behavior against to which judge our teacher selection procedures. Such correlates-type estimates may involve (a) behavior or products from simulated situations (e.g., performance situations, simulating those situations in which the criterion behavior occurs) or, (b) even presentation of graphic and/or verbal descriptions of situations involving the criterion behavior.

As I come to the conclusion of my remarks I feel a strong sense of inadequacy; of having bitten off more than I can chew. As is the case with most of you present, I have given a great deal of thought to the problem of the criterion, particularly as it relates to teacher behavior and to the problem of validity study of teacher selection devices. I find it easy to identify and recognize many of the problems and difficulties with which we are faced in trying to develop satisfactory descriptions of the criterion behavior of teachers and techniques which will yield valid estimates of the criterion behavior involved in teaching. I recognize the sources of bias in the description of criterion behavior and the conditions making for invalidity of the estimates yielded by different methods

of assessing criterion behavior. But I am admittedly frustrated by the difficulties involved in obtaining criterion data which are, on one hand, inclusive and complete and, on the other, exclusive and free of contamination. I know it is not easy to lick these problems, particularly when we must frequently conduct validity studies in situations where we have been using certain teacher selection devices that were selected on a priori basis without the benefit of guidance of adequate criterion descriptions. And now, after the fact, we are faced with the problem of providing procedures that will yield estimates of criterion descriptions against which to test our selection data. I do not think the situation is an impossible one, but I cannot help but recognize, as I think most of us must, that we are faced with practical considerations which force us to compromise and employ make-shift methods that preclude the carrying out of validity studies of the quality we would like.